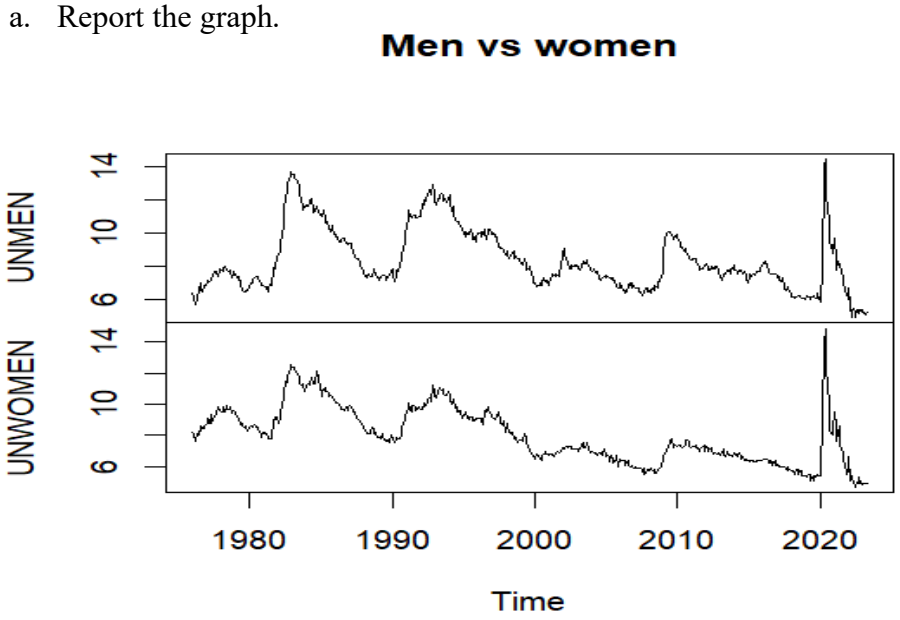
Udit Lal
2/17/2024
ECON 4385

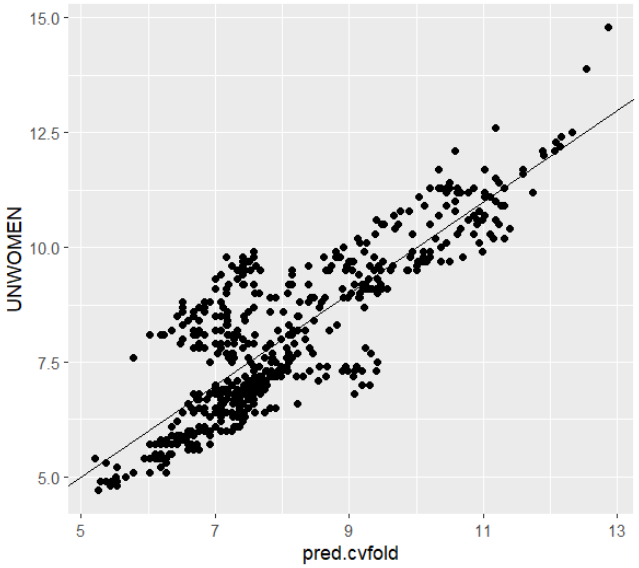
HW#4

1. Adjust the code to use the Canadian data set in line 40. Adjust the start and end dates to use the start and end dates for the Canadian data set (lines 46-62). Deliverable: changes to your code. No deliverable for the write up for this question.
2. Graph the Canadian unemployment rate for men versus the unemployment rate for women in a single graph. Include the graph you obtained in your answer (screenshots are not acceptable). Note: If you uploaded the data correctly, the code will generate this graph for you, but you have to export it and include it in your file.
    a. Report the graph.



    b. What is the first observation included in your data set? What is the last observation included in your data set.
        i. The first datapoint is for January 1, 1976
        The last datapoint is for April 1, 2023
    c. Comment on the graph's features. In particular, discuss whether male and female unemployment rates usually move in sync? When were they elevated? When were they low?
        i. The unemployment rate for men and women move very similarly. They have the same pattern just to different degrees.
    d. Can you make an argument based on an economic or political theory that there is a link between these two unemployment rates? I am not asking for a deep economic theory or an essay here, a short 1-2 sentence answer based on principles of economics is fine.
        i.

3. Regress the unemployment rate for women on a constant and the unemployment rate for men
    a. Report the coefficient estimates, the r-squared, and RMSE for the entire sample. The code will automatically generate these results for you, but you will need to report them yourself. Screen shots are not acceptable.
        i. The β coefficient for the unemployment rate of men on the unemployment rate of women is: 0.82
        r-squared: 0.71
        RMSE: 0.96
    b. Discuss the coefficient on UNMEN. How can you interpret this coefficient in English in 1-2 sentences?
        i. For every unit (%) that the unemployment rate increases, the unemployment rate for women increases by 0.82 units (%)
4. Using a 5 way fold predict the unemployment rate for women using the unemployment rate for men. The code is set up to predict using a 5-way-fold and you do not need to make any changes for this part.
    a. Report the r-squared and the RMSE for the 5-way fold.
        i.
    b. How do they compare to the r-squared and the RMSE for the entire sample? Are they much larger, much smaller, or roughly comparable?
5. Report and discuss the graph with predicted unemployment rates vs observed unemployment rates for women for a 5-way-fold. The code is set up to predict using a 5-way-fold and you do not need to make any changes for this part.
    a. Report the graph with the predicted versus the actual unemployment rates. Screenshots are not acceptable.
        i.



    b. Is there any visual evidence that the model systematically over or under-predicts?

    i. The model passes the eye test when it comes to accuracy. It looks like there is an about equal number of points above the line as below.

c. Is there any visual evidence that the model is better at predicting low rates or high rates? There are two ways evaluate this: you can use the 45-degree line and the actual versus predicted values, and check if the points are systematically below or systematically above the 45-degree line when unemployment is low or when it is high. Alternatively, you can use a Gain Curve Plot and the 45-degree line. If the predicted line is below the 45-degree line for low values (first 10-30%= fractions in order 0.1 to 0.3), then the model is not appropriate for predicting low unemployment rates. If the predicted line is below the 45-degree line for high values (highest 70-100%= fractions in order 0.7 to 1), then the model is not appropriate for predicting high unemployment rates.

    i. The model is best when used for medium unemployment rates. The model sits at the top of the low estimates giving a higher number than most datapoints and vice versa for high unemployment rates.

6. Repeat questions 4 and 5 using a 10-way fold. Do the answers look different from the answers to questions 4 and 5?

    a. Using a higher number of folds did not really improve the model as the r-squared and the RMSE were about the same in both models